

1645

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/292,437

DATE: 01/04/2001  
 TIME: 08:34:49

Input Set : A:\#123572 v1 - REVISED SEQUENCE LISTING.txt  
 Output Set: N:\CRF3\01042001\I292437.raw

RECEIVED  
 JAN 11 2001  
 TECH CENTER 1600/2900

ENTERED

4. <110> APPLICANT: Olaf Schneewind  
 5 Sarkis Mazmanian  
 6 Gwen Liu  
 7 Hung Ton-That  
 9 <120> TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE  
 12 <130> FILE REFERENCE: 510015.213  
 W--> 13 <140> CURRENT APPLICATION NUMBER: US/09/292,437  
 14 <141> CURRENT FILING DATE: 1999-04-15  
 16 <160> NUMBER OF SEQ ID NOS: 36  
 18 <170> SOFTWARE: FastSEQ for Windows Version 3.0  
 20 <210> SEQ ID NO: 1  
 21 <211> LENGTH: 5  
 22 <212> TYPE: PRF  
 23 <213> ORGANISM: Bacteria  
 25 <220> FEATURE:  
 26 <221> NAME/KEY: Unsure  
 27 <222> LOCATION: 3  
 28 <223> OTHER INFORMATION: Any naturally-occurring amino acid  
 30 <400> SEQUENCE: 1  
 W--> 31 Leu Pro Xaa Thr Gly  
 32 1 5  
 34 <210> SEQ ID NO: 2  
 35 <211> LENGTH: 621  
 36 <212> TYPE: DNA  
 37 <213> ORGANISM: Staphylococcus aureus  
 39 <400> SEQUENCE: 2  
 40 atg aaa aaa tgg aca aat cga tta atg aca atc gct ggt gtg gta ctt 48  
 41 Met Lys Lys Trp Thr Asn Arg Leu Met Thr Ile Ala Gly Val Val Leu  
 42 1 5 10 15  
 44 atc cta gtg gca gca tat ttg ttt gct aaa cca cat atc gat aat tat 96  
 45 Ile Leu Val Ala Ala Tyr Leu Phe Ala Lys Pro His Ile Asp Asn Tyr  
 46 20 25 30  
 48 ctt cac gat aaa gat aaa gat gaa aag att gaa caa tat gat aaa aat 144  
 49 Leu His Asp Lys Asp Lys Asp Glu Lys Ile Glu Gln Tyr Asp Lys Asn  
 50 35 40 45  
 52 gta aaa gaa cag gcg agt aaa gal aaa aag caa gct aaa cct caa 192  
 53 Val Lys Glu Gln Ala Ser Lys Asp Lys Lys Glu Gln Ala Lys Pro Gln  
 54 50 55 60  
 56 att ccg aaa gat aaa tcg aaa gtg gca ggc tat att gaa att cca gat 240  
 57 Ile Pro Lys Asp Lys Ser Lys Val Ala Gly Tyr Ile Glu Ile Pro Asp  
 58 65 70 75 80  
 60 gct gat att aaa gaa cca gta tat cca gga cca gca aca cct gaa caa 288  
 61 Ala Asp Ile Lys Glu Pro Val Tyr Pro Gly Pro Ala Thr Pro Glu Gln  
 62 85 90 95  
 64 tta aat aga ggt gta agc ttt gca gaa gaa aat gaa tca cta gat gat 336  
 65 Leu Asn Arg Gly Val Ser Phe Ala Glu Glu Asn Glu Ser Leu Asp Asp  
 66 100 105 110

BEST AVAILABLE COPY

## RAW SEQUENCE LISTING

DATE: 01/04/2001

PATENT APPLICATION: US/09/292,437

TIME: 08:34:49

Input Set : A:\#123572 v1 - REVISED SEQUENCE LISTING.txt

Output Set: N:\CRF3\01042001\I292437.raw

```

68  caa aat att tca att gca gga cac act ttc att gac cgt ccg aac tat      384
69  Gln Asn Ile Ser Ile Ala Gly His Thr Phe Ile Asp Arg Pro Asn Tyr
70          115          120          125
72  caa ttt aca aat ctt aaa gca gcc aaa aaa ggt agt atg gtg tac ttt      432
73  Gln Phe Thr Asn Leu Lys Ala Ala Lys Lys Gly Ser Met Val Tyr Phe
74          130          135          140
76  aaa gtt ggt aat gaa aca cgt aaq tat aaa atg aca aqt ata aga gat      480
77  Lys Val Gly Asn Glu Thr Arg Lys Tyr Lys Met Thr Ser Ile Arg Asp
78  145          150          155          160
80  gtt aag cct aca gat gta gga gtt cta gat gaa caa aaa ggt aaa gat      528
81  Val Lys Pro Thr Asp Val Gly Val Leu Asp Glu Gln Lys Gly Lys Asp
82          165          170          175
84  aaa caa tta aca tta att act tgt gat tat tuc aat gaa aag aca ggc      576
85  Lys Gln Leu Thr Leu Ile Thr Cys Asp Asp Tyr Asn Glu Lys Thr Gly
86          180          185          190
88  gtt tgg gaa aaa cgt aaa atc ttt gta gct aca gaa gtc aaa taa      621
89  Val Trp Glu Lys Arg Lys Ile Phe Val Ala Thr Glu Val Lys. *
90          195          200          205

93 <210> SEQ ID NO: 3
94 <211> LENGTH: 206
95 <212> TYPE: PRT
96 <213> ORGANISM: Staphylococcus aureus
98 <400> SEQUENCE: 3
99  Met Lys Lys Trp Thr Asn Arg Leu Met Thr Ile Ala Gly Val Val Leu
100  1          5          10          15
101  Ile Leu Val Ala Ala Tyr Leu Phe Ala Lys Pro His Ile Asp Asn Tyr
102          20          25          30
103  Leu His Asp Lys Asp Lys Asp Glu Lys Ile Glu Gln Tyr Asp Lys Asn
104          35          40          45
105  Val Lys Glu Gln Ala Ser Lys Asp Lys Lys Gln Gln Ala Lys Pro Gln
106          50          55          60
107  Ile Pro Lys Asp Lys Ser Lys Val Ala Gly Tyr Ile Glu Ile Pro Asp
108          65          70          75          80
109  Ala Asp Ile Lys Glu Pro Val Tyr Pro Gly Pro Ala Thr Pro Glu Gln
110          85          90          95
111  Leu Asn Arg Gly Val Ser Phe Ala Glu Glu Asn Glu Ser Leu Asp Asp
112          100          105          110
113  Gln Asn Ile Ser Ile Ala Gly His Thr Phe Ile Asp Arg Pro Asn Tyr
114          115          120          125
115  Gln Phe Thr Asn Leu Lys Ala Ala Lys Lys Gly Ser Met Val Tyr Phe
116          130          135          140
117  Lys Val Gly Asn Glu Thr Arg Lys Tyr Lys Met Thr Ser Ile Arg Asp
118          145          150          155          160
119  Val Lys Pro Thr Asp Val Gly Val Leu Asp Glu Gln Lys Gly Lys Asp
120          165          170          175
121  Lys Gln Leu Thr Leu Ile Thr Cys Asp Asp Tyr Asn Glu Lys Thr Gly
122          180          185          190
123  Val Trp Glu Lys Arg Lys Ile Phe Val Ala Thr Glu Val Lys
124          195          200          205

```

BEST AVAILABLE COPY

RAW SEQUENCE LISTING                      DATE: 01/04/2001  
 PATENT APPLICATION: US/09/292,437                      TIME: 08:34:49

Input Set : A:\#123572 v1 - REVISED SEQUENCE LISTING.txt  
 Output Set: N:\CRF3\01042001\I292437.raw

```

126 <210> SEQ ID NO: 4
127 <211> LENGTH: 227
128 <212> TYPE: PRT
129 <213> ORGANISM: Streptococcus pyogenes
131 <400> SEQUENCE: 4
132 Met Glu Glu Val Trp Glu Lys Ala Lys Ala Tyr Asn Ala Arg Leu Gly
133 1 5 10 15
134 Thr Gln Pro Val Pro Asp Ala Phe Ser Phe Arg Asp Gly Ile His Asp
135 20 25 30
136 Lys Asn Tyr Glu Ser Leu Leu Gln Ile Glu Asn Asn Asp Ile Met Gly
137 35 40 45
138 Tyr Val Glu Val Pro Ser Ile Lys Val Thr Leu Pro Ile Tyr His Tyr
139 50 55 60
140 Thr Thr Asp Glu Val Leu Thr Lys Gly Ala Gly His Leu Phe Gly Ser
141 65 70 75 80
142 Ala Leu Pro Val Gly Asp Gly Thr His Thr Val Ile Ser Ala His
143 85 90 95
144 Arg Gly Leu Pro Ser Ala Glu Met Phe Thr Asn Leu Asn Leu Val Lys
145 100 105 110
146 Lys Gly Asp Thr Phe Tyr Phe Arg Val Leu Asn Lys Val Leu Ala Tyr
147 115 120 125
148 Lys Val Asp Gln Ile Leu Thr Val Glu Pro Asp Gln Val Thr Ser Leu
149 130 135 140
150 Ser Gly Val Met Gly Lys Asp Tyr Ala Thr Leu Val Thr Cys Thr Pro
151 145 150 155 160
152 Tyr Gly Val Asn Thr Lys Arg Leu Leu Val Arg Gly His Arg Ile Ala
153 165 170 175
154 Tyr His Tyr Lys Lys Tyr Gln Gln Ala Lys Lys Ala Met Lys Leu Val
155 180 185 190
156 Asp Lys Ser Arg Met Trp Ala Glu Val Val Cys Ala Ala Phe Gly Val
157 195 200 205
158 Val Ile Ala Ile Ile Leu Val Phe Met Tyr Ser Arg Val Ser Ala Lys
159 210 215 220
160 Lys Ser Lys
161 225
163 <210> SEQ ID NO: 5
164 <211> LENGTH: 365
165 <212> TYPE: PRT
166 <213> ORGANISM: Actinomyces naeslundii
168 <400> SEQUENCE: 5
169 Met Gly Leu Leu Thr Tyr Pro Thr Ala Ala Ser Trp Val Ser Gln Tyr
170 1 5 10 15
171 Asn Gln Ser Lys Val Thr Ala Asp Tyr Ser Ala Gln Val Asp Gly Ala
172 20 25 30
173 Arg Pro Asp Ala Lys Thr Gln Val Glu Gln Ala His Ala Tyr Asn Asp
174 35 40 45
175 Ala Leu Ser Ala Gly Ala Val Leu Glu Ala Asn Asn His Val Pro Thr
176 50 55 60
177 Gly Ala Gly Ser Ser Lys Asp Ser Ser Leu Gln Tyr Ala Asn Ile Leu

```

**BEST AVAILABLE COPY**

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/292,437

DATE: 01/04/2001

TIME: 08:34:49

Input Set : A:\#123572 v1 - REVISED SEQUENCE LISTING.txt

Output Set: N:\CRF3\01042001\I292437.raw

178 65 70 75 80  
 179 Lys Ala Asn Asn Glu Gly Leu Met Ala Arg Leu Lys Ile Pro Ser Ile  
 180 85 90 95  
 181 Ser Leu Asp Leu Pro Val Tyr His Gly Thr Ala Asp Asp Thr Leu Leu  
 182 100 105 110  
 183 Lys Gly Leu Gly His Leu Glu Gly Thr Ser Leu Pro Val Gly Gly Glu  
 184 115 120 125  
 185 Gly Thr Arg Ser Val Ile Thr Gly His Arg Gly Leu Ala Glu Ala Thr  
 186 130 135 140  
 187 Met Phe Thr Asn Leu Asp Lys Val Lys Thr Gly Asp Ser Leu Ile Val  
 188 145 150 155 160  
 189 Glu Val Phe Gly Glu Val Leu Thr Tyr Arg Val Thr Ser Thr Lys Val  
 190 165 170 175  
 191 Val Glu Pro Glu Glu Thr Glu Ala Leu Arg Val Glu Glu Gly Lys Asp  
 192 180 185 190  
 193 Leu Leu Thr Leu Val Thr Cys Thr Pro Leu Gly Ile Asn Thr His Arg  
 194 195 200 205  
 195 Ile Leu Leu Thr Gly Glu Arg Ile Tyr Pro Thr Pro Ala Lys Asp Leu  
 196 210 215 220  
 197 Ala Ala Ala Gly Lys Arg Pro Asp Val Pro His Phe Pro Trp Trp Ala  
 198 225 230 235 240  
 199 Val Gly Leu Ala Ala Gly Leu Ile Val Val Gly Leu Tyr Leu Trp Arg  
 200 245 250 255  
 201 Ser Gly Tyr Ala Ala Ala Arg Ala Lys Glu Arg Ala Leu Ala Arg Ala  
 202 260 265 270  
 203 Arg Ala Ala Gln Glu Glu Pro Gln Pro Gln Thr Trp Ala Glu Gln Met  
 204 275 280 285  
 205 Arg Ile Trp Met Asp Asp Asp Ala Gly Val Glu Pro Gln Arg Trp Phe  
 206 290 295 300  
 207 Thr Asp Leu Pro Val Pro Pro Gln Pro Ser Glu Met Glu Asn Leu Ala  
 208 305 310 315 320  
 209 Leu Leu Glu Glu Ile Ala Ser Leu Ser Ala Pro Ser Gly Arg Trp Asp  
 210 325 330 335  
 211 Asp Gln Glu Leu Ile Asp Thr Ala Glu Ile Pro Val Leu Asp Ala Thr  
 212 340 345 350  
 213 Arg Pro Ser Ala Gly Thr Ser Gly Arg Thr His Arg Leu  
 214 355 360 365  
 216 <210> SEQ ID NO: 6  
 217 <211> LENGTH: 284  
 218 <212> TYPE: PRT  
 219 <213> ORGANISM: Enterococcus faecalis  
 221 <400> SEQUENCE: 6  
 222 Met Lys Ser Lys Lys Arg Arg Ile Ile Asp Gly Phe Met Ile Leu  
 223 1 5 10 15  
 224 Leu Leu Ile Ile Gly Ile Gly Ala Phe Ala Tyr Pro Phe Val Ser Asp  
 225 20 25 30  
 226 Ala Leu Asn Asn Tyr Leu Asp Gln Gln Ile Ile Ala His Tyr Gln Ala  
 227 35 40 45  
 228 Lys Ala Ser Gln Glu Asn Thr Lys Glu Met Ala Glu Leu Gln Glu Lys

RECEIVED  
 JAN 11 2001  
 TECH CENTER 1600/2900

BEST AVAILABLE COPY

## RAW SEQUENCE LISTING

DATE: 01/04/2001

PATENT APPLICATION: US/09/292,437

TIME: 08:34:50

Input Set : A:\#123572 v1 - REVISED SEQUENCE LISTING.txt

Output Set: N:\CRF3\01042001\I292437.raw

```

229      50      55      60
230 Met Glu Lys Lys Asn Gln Glu Leu Ala Lys Lys Gly Ser Asn Pro Gly
231 65      70      75      80
232 Leu Asp Pro Phe Ser Glu Thr Gln Lys Thr Thr Lys Lys Pro Asp Lys
233      85      90      95
234 Ser Tyr Phe Glu Ser His Thr Ile Gly Val Leu Thr Ile Pro Lys Ile
235      100      105      110
236 Asn Val Arg Leu Pro Ile Phe Asp Lys Thr Asn Ala Leu Leu Leu Glu
237      115      120      125
238 Lys Gly Ser Ser Leu Leu Glu Gly Thr Ser Tyr Pro Thr Gly Gly Thr
239      130      135      140
240 Asn Thr His Ala Val Ile Ser Gly His Arg Gly Leu Pro Gln Ala Lys
241 145      150      155      160
242 Leu Phe Thr Asp Leu Pro Glu Leu Lys Lys Gly Asp Glu Phe Tyr Ile
243      165      170      175
244 Glu Val Asn Gly Lys Thr Leu Ala Tyr Gln Val Asp Gln Ile Lys Thr
245      180      185      190
246 Val Glu Pro Thr Asp Thr Lys Asp Leu His Ile Glu Ser Gly Gln Asp
247      195      200      205
248 Leu Val Thr Leu Leu Thr Cys Thr Pro Tyr Met Ile Asn Ser His Arg
249      210      215      220
250 Leu Leu Val Arg Gly His Arg Ile Pro Tyr Gln Pro Glu Lys Ala Ala
251 225      230      235      240
252 Ala Gly Met Lys Lys Val Ala Gln Gln Gln Asn Leu Leu Leu Trp Thr
253      245      250      255
254 Leu Leu Leu Ile Ala Cys Ala Leu Ile Ile Ser Gly Phe Ile Ile Trp
255      260      265      270
256 Tyr Lys Arg Arg Lys Lys Thr Thr Arg Lys Pro Lys
257      275      280
259 <210> SEQ ID NO: 7
260 <211> LENGTH: 246
261 <212> TYPE: PRT
262 <213> ORGANISM: Streptococcus mutans
264 <400> SEQUENCE: 7
265 Met Lys Lys Glu Arg Gln Ser Arg Lys Lys Arg Ser Phe Leu Arg Thr
266 1      5      10      15
267 Phe Leu Pro Ile Leu Leu Leu Val Ile Gly Leu Ala Leu Ile Phe Asn
268      20      25      30
269 Thr Pro Ile Arg Asn Ala Leu Ile Ala Trp Asn Thr Asn Arg Tyr Gln
270      35      40      45
271 Val Ser Asn Val Ser Lys Lys Asp Ile Glu His Asn Lys Ala Ala His
272      50      55      60
273 Ser Ser Phe Asp Phe Lys Lys Val Glu Ser Ile Ser Thr Gln Ser Val
274      65      70      75      80
275 Leu Ala Ala Gln Met Ala Ala Gln Lys Leu Pro Val Ile Gly Gly Ile
276      85      90      95
277 Ala Ile Pro Asp Leu Lys Ile Asn Leu Pro Ile Phe Lys Gly Leu Asp
278      100      105      110
279 Asn Val Gly Leu Thr Tyr Gly Ala Gly Thr Met Lys Asn Asp Gln Val

```

BEST AVAILABLE COPY

VERIFICATION SUMMARY                      DATE: 01/04/2001  
PATENT APPLICATION:    US/09/292,437              TIME: 08:34:51

Input Set : A:\#123572 v1 - REVISED SEQUENCE LISTING.txt  
Output Set: N:\CRF3\01042001\I292437.raw

L:13 M:283 W: Missing Blank Line separator, <140> field identifier  
L:31 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1

**BEST AVAILABLE COPY**